DOCUMENT RESUME

ED 052 866 RC 005 457

AUTHOR	Hines, Brainard W.
TITLE	Analysis of Intelligence Scores.
INSTITUTION	Appalachia Educational Lab., Charleston, W. Va.
REPORT NO	TR-2
PUB DATE	Feb 71
NOTE	16p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Culturally Disadvantaged, *Early Childhood Education, *Evaluation, Home Instruction, *Intelligence Tests, Sex Differences, Statistical Analysis, Subprofessionals, Tables (Data), *Televised Instruction
IDENTIFIERS	*Appalachia, West Virginia

ABSTRACT

In September 1969, a random sample of 160 3- and 4-year-olds was selected from a larger group of individuals in 3 Early Childhood Education Program treatment groups (television only; television + home visitor; and television + home visitor + mobile classroom), along with a no-treatment comparison group, for the purpose of analyzing intelligence scores and congruence with program objectives. These children were tested in June and September 1970 by individuals trained by the Appalachia Educational Laboratory but not otherwise involved in the program. In this way, it was hoped that any examiner bias would be minimized and that which remained would be a constant factor throughout all treatment groups. Data from the 4 groups were analyzed by a 3-way analysis of variance (ANOVA) and an analysis of covariance (ANCOVA) procedure. The ANOVA procedure involved 4 treatment levels, 2 age levels, and 2 levels of sex; the ANCOVA used the "Peabody Picture Vocabulary Test" raw score and chronological age as covariates. The BMD16V general linear hypothesis program for unbalanced design was used to run the ANOVA and ANCOVA for each variable. There was no need to test the homogeneity of variances since the program adjusts for any lack of homogeneity. In the document, graphs of significant interaction effects are presented for each subtest where such effects occurred, along with results of Scheffe post hoc comparisons. In the analysis of variance table for each subtest, the sum of squares column is replaced by a list of eta squared calculations. (LS)

ED052866

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ANALYSIS OF INTELLIGENCE SCORES

Technical Report No. 2

Division of Research and Evaluation
Appalachia Educational Laboratory
Charleston, West Virginia

February, 1971

RC005457

TABLE OF CONTENTS

PAGE

DESCRIPTION OF THE PPVT 1

METHOD 3

SUMMARY OF FINDINGS 4

SUMMARY AND CONCLUSIONS 13

LIST OF TABLES

TABLE

2-1 PPVT RAW SCORE MEANS, STANDARD DEVIATIONS, AND NUMBER OF SUBJECTS BY AGE, SEX, AND TREATMENTS 4

2-2 PPVT RAW SCORE MEANS, STANDARD DEVIATIONS, AND SAMPLE SIZE BY TREATMENT GROUPS 5

2-3 ANALYSIS OF VARIANCE TABLE FOR PPVT RAW SCORES 6

2-4 PPVT MENTAL AGE MEANS, STANDARD DEVIATIONS, AND NUMBER OF SUBJECTS BY AGE, SEX, AND TREATMENTS 8

2-5 MENTAL AGE (IN MONTHS) MEANS, STANDARD DEVIATIONS, AND SAMPLE SIZES BY TREATMENT GROUPS 8

2-6 ANALYSIS OF VARIANCE SUMMARY TABLE FOR MENTAL AGE SCORES 10

2-7 PPVT IQ SCORE MEANS, STANDARD DEVIATIONS, AND NUMBER OF SUBJECTS BY AGE, SEX, AND TREATMENTS 11

2-8 IQ SCORE MEANS, STANDARD DEVIATIONS, AND SAMPLE SIZES FOR FOUR TREATMENTS 11

2-9 SUMMARY OF ANALYSIS OF VARIANCE FOR PPVT IQ SCORES 12

LIST OF FIGURES

FIGURE		PAGE
2-1	PPVT RAW SCORE MEANS AND NATIONAL NORMS BY TREATMENT GROUPS	5
2-2	MEAN MENTAL AGE (IN MONTHS) FOR FOUR TREATMENT GROUPS	9
2-3	IQ SCORE MEANS FOR FOUR TREATMENT GROUPS	12

ANALYSIS OF INTELLIGENCE SCORES*

Because of the varied definitions of intelligence, many of which are dependent on a particular psychometric technique or conceptual theory, the AEL Early Childhood evaluation uses an operational definition of intelligence. Rather than dealing with a specific theory, the major AEL evaluation instrument of intellectual growth, the Peabody Picture Vocabulary Test, is the "operation" which provides our definition of verbal IQ and verbal intelligence.

Although this procedure avoids the difficulty of conflicting theories, it runs the risk of a poor fit between program definitions and effects on one hand, and the sensitivity of instruments on the other. That is, the meaning of intelligence which is implied in the ECE curriculum may differ from that which was utilized in the development of the evaluation instrument.

For the above reasons, this report will be concerned with the theory underlying the PPVT as well as its congruence with the ECE program objectives. In addition, data gathered in June and September of 1970 will be presented along with a summary of the analyses performed on the raw scores from several treatment groups, as well as on the derived scores (mental age and IQ) which were obtained from the raw scores.

DESCRIPTION OF THE PPVT

The Peabody Picture Vocabulary Test consists of a series of 150 plates, each of which is comprised of four separate illustrations. One of the four illustrations on each plate corresponds to a key word chosen from Webster's New Collegiate Dictionary (G & C Merriam, 1953), and is included in the body of the test. The examiner begins at a basal level in the test and pronounces each word on the list, showing the child the particular plate which contains

---

*This report was prepared by Brainard W. Hines of the Research and Evaluation Division.

the illustration of the word just pronounced. The child responds by pointing at the correct illustration, and the examiner records the response as correct or incorrect. After a series of six incorrect answers in eight responses, testing is discontinued. The total raw score of correct responses for the test is calculated, and a mental age (M.A.) is derived from the total score. In addition, raw score and chronological age are used to derive a deviation IQ-score, utilizing a mean of 100 and a standard deviation of fifteen.

Several comments on the general format and theoretical basis of the instrument are appropriate at this point. First, it is obvious that the test depends solely on the child's verbal ability, and on a narrow range of that particular factor. Insofar as the PPVT's vocabulary-type format reflects the same general factor as do longer tests, such as the Stanford-Binet or the Wechsler scales, it should have a similar predictive ability for future school success. It is likely that the vocabulary level measured by the PPVT correlates approximately .70 with total verbal ability, which in turn correlates about .50 with later school success.[1]

Finally, the nonverbal response which is required from the child is easily influenced by the examiner's biases and resulting cues. This possibility is inevitable in any instrument which is capable of being quickly administered to children of preschool age. To minimize differential bias on the child's pattern of responding, the testers should be relatively naive in the area of program effort to be evaluated, and should be trained to be as objective as possible when administering the instrument.

The advantages of the instrument outweigh the previous considerations. It is easily administered, reliable, provides alternate forms, and correlates fairly highly with more time consuming instruments. The verbal functions which

---

[1] Expanded Manual for the Peabody Picture Vocabulary Test, (Lloyd M. Dunn), American Guidance Service, Inc., Minneapolis, Minnesota., pp. 35-40.

it measures are closely related to many of the ECE program goals. It provides an estimate of verbal ability and verbal intelligence but is not designed to provide approximation of Spearman's 'g' factor of intelligence, or general intellectual ability.

It should further be noted that the IQ derived from raw score and chronological age is a deviation score and not an "intelligence quotient". That is, it is a scaled score which represents the mean of the normative group with a score of 100 and a standard deviation of fifteen points. For this reason, the IQ scores which it produces are not directly comparable with those which are mathematically derived by the mental age/chronological age formula.

METHOD

A sample of 160 children, aged three and four, in September 1969, was randomly selected from a larger group of individuals in three treatment groups and one control group. A detailed description of this process is presented in the discussion of sampling procedures.

These children were tested in June and September 1970 by a group of individuals trained by AEL, but not otherwise involved in the program. In this way it was hoped that any examiner bias would be minimized and that which remained would be a constant factor throughout all treatment groups.

Data from these four groups were analyzed by means of both a three-way analysis of variance (ANOVA) and an analysis of covariance (ANCOVA) procedure. The ANOVA procedure involved four levels of treatment, two levels of age and two levels of sex, while the ANCOVA used PPVT raw score and chronological age as covariates.

The BMDX64 general linear hypothesis program[2] for unbalanced design was

[2]W.J.Dixon, Editor, Biomedical Computer Program, University of California Press, 1970. The analysis was performed at the University of Michigan Computing Center.

used to run the ANOVA and ANCOVA for each variable. With this program there was no need to test the homogeneity of variances since it adjusts for any lack of homogeneity.

In addition, graphs of significant interaction effects will be presented for each subtest where such effects occurred, along with the results of Scheffe post hoc comparisons. In the analysis of variance table for each subtest, the sum of squares column will be replaced by a list of eta squared calculations. Eta squared is the proportion of variance accounted for by each source and is determined by dividing each sum of squares by the total of the sums of squares.

SUMMARY OF FINDINGS

PPVT Raw Scores

The raw scores on the PPVT, which consist of the total number of correct responses given throughout the test, are recorded for each treatment group by age and sex cell below in Table 2-1.

TABLE 2-1

PPVT RAW SCORE MEANS, STANDARD DEVIATION,
AND NUMBER OF SUBJECTS BY AGE, SEX, AND TREATMENTS

Age	Sex	Package	TV-HV	TV only	Control
3	Male	$\bar{x}$ = 42.63 SD = 6.55 N = 8	$\bar{x}$ = 38.56 SD = 13.33 N = 9	$\bar{x}$ = 40.46 SD = 11.60 N = 13	$\bar{x}$ = 42.23 SD = 10.51 N = 13
	Female	$\bar{x}$ = 42.63 SD = 9.88 N = 8	$\bar{x}$ = 42.80 SD = 8.05 N = 10	$\bar{x}$ = 32.90 SD = 14.05 N = 10	$\bar{x}$ = 37.31 SD = 7.72 N = 13
4	Male	$\bar{x}$ = 50.38 SD = 8.23 N = 13	$\bar{x}$ = 51.13 SD = 6.22 N = 8	$\bar{x}$ = 39.88 SD = 18.16 N = 8	$\bar{x}$ = 45.89 SD = 8.37 N = 9
	Female	$\bar{x}$ = 47.09 SD = 9.43 N = 11	$\bar{x}$ = 48.20 SD = 10.35 N = 10	$\bar{x}$ = 44.31 SD = 11.95 N = 13	$\bar{x}$ = 49.70 SD = 6.88 N = 10

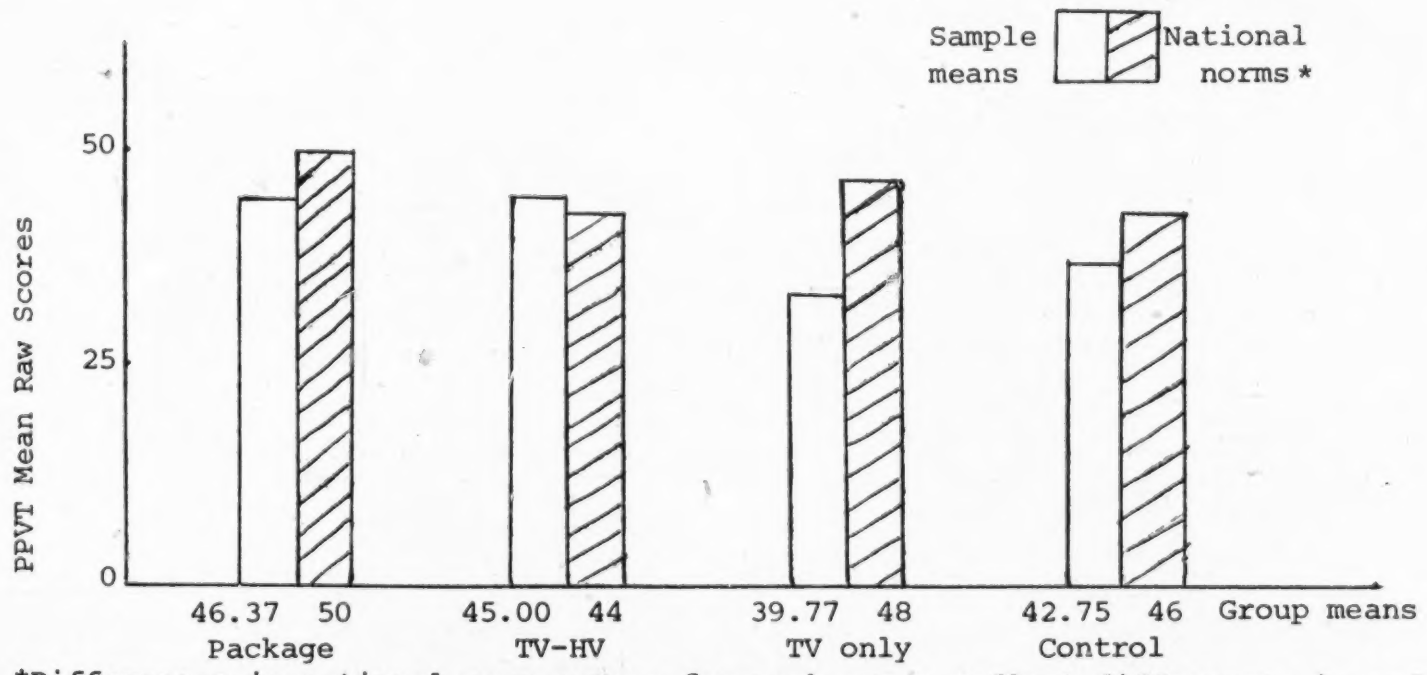
These raw scores are difficult to interpret, even between groups, since no normative data are included, and we would expect raw scores to vary with mean age for each group.

Therefore, each overall treatment group mean is presented in Figure 2-1, and mean raw scores for each normative age comparison group are given in Table 2-2.

TABLE 2-2

PPVT RAW SCORE MEANS, STANDARD DEVIATIONS,
AND SAMPLE SIZE BY TREATMENT GROUPS

	Package	TV-HV	TV-only	Control
$\bar{X}$	46.37	45.0	39.77	42.75
SD	8.95	10.01	13.78	9.79
N	40	37	44	45



*Differences in national norm scores for each group reflect differences in age

FIGURE 2-1

PPVT RAW SCORE MEANS AND NATIONAL
NORMS BY TREATMENT GROUPS

The graphs in Figure 2-1 reveal several interesting trends. First, all four groups scored below the normative sample in overall vocabulary level. Second, the tendency for the TV only group to score below the other treatments, which is evident throughout most of the testing battery, is also present in this case. And, finally, the two groups which received visits from paraprofessionals tended to score above those groups who did not receive such visits.

The analysis of variance procedure which was performed on the above raw scores produced the following results, summarized in Table 2-3.

TABLE 2-3

ANALYSIS OF VARIANCE TABLE FOR PPVT RAW SCORES

Source	M^2*	df	Mean Square	F	P
I(trt)	.048	3	329.8072186	3.00	$P < .05$
J(sex)	.001	1	24.96025017	0.23	
K(age)	.105	1	2142.289759	19.52	$P < .0005$
IJ-INT	.002	3	10.63576751	0.10	
IK-INT	.005	3	33.39158528	0.30	
JK-INT	.003	1	64.67400307	0.59	
IJKINT	.031	3	213.3936408	1.94	
Error		150	109.7637244		

*Eta squared (M^2) is the proportion of variance accounted for by each source and is determined by dividing each sums of squares by the total sums of squares. A convenient reference is: Hays, William L., Statistics Holt, Rinehart and Winston, 1963, p. 546-548.

The marked significance of the main effect of age was as expected from any measure. Consequently, the main effect of treatment is even more striking in that the TV only sample scored somewhat below the comparison group even though they were slightly older in chronological age. That is, since age and PPVT raw score seem to be highly correlated, it is suprising to find a group with a higher mean age producing a lower mean raw score than a given comparison group.

A post-hoc comparison utilising the Scheffe test indicated that the TV only group was scoring significantly lower than the Package group, and that this difference was accounting for the treatment effect which was apparent in the analysis of variance.

Keeping these facts in mind, the probability of a main effect of treatment leads to several inferences about program effectiveness. First, the paraprofessional home visitor seems to contribute to the level of learning measured by the PPVT. The two groups which receive visits from the paraprofessional show elevated means when compared to those which view only the television program or are not exposed to any of the program elements. Second, the relatively depressed scores which are apparent for the TV only group may well be indicative of a lower level of socio-economic status for this part of the sample.

One hundred percent of the TV only group lived in a rural section of the county as opposed to sixty to seventy percent for the other two groups.

PPVT Mental Age

The MA score which appears on the Peabody is a derived score, based on the average age of the subsample within the normative group which was able to respond correctly to a given total of test items. In this way, a mental age score of four years three months indicates that in the normative sample, a majority of children of this mean age were able to obtain a specific raw score, which in this case would be 44.

Table 3-4 lists mean mental ages for Form B of the PPVT for each age by sex cell within the four treatment groups.

TABLE 2-4

PPVT MENTAL AGE MEANS, STANDARD DEVIATIONS AND NUMBER OF SUBJECTS BY AGE, SEX, AND TREATMENTS

Age	Sex	Package	TV-HV	TV only	Control
3	Male	$\bar{x}$ = 51.13 SD = 8.98 N = 8	$\bar{x}$ = 48.11 SD = 15.85 N = 9	$\bar{x}$ = 50.46 SD = 15.02 N = 13	$\bar{x}$ = 50.85 SD = 13.08 N = 13
	Female	$\bar{x}$ = 52.88 SD = 14.58 N = 8	$\bar{x}$ = 51.80 SD = 12.58 N = 10	$\bar{x}$ = 41.80 SD = 12.59 N = 10	$\bar{x}$ = 40.69 SD = 14.06 N = 13
4	Male	$\bar{x}$ = 64.38 SD = 14.15 N = 13	$\bar{x}$ = 65.13 SD = 10.93 N = 8	$\bar{x}$ = 51.75 SD = 19.85 N = 8	$\bar{x}$ = 57.00 SD = 11.63 N = 9
	Female	$\bar{x}$ = 59.10 SD = 16.37 N = 11	$\bar{x}$ = 59.80 SD = 16.89 N = 10	$\bar{x}$ = 56.00 SD = 15.47 N = 13	$\bar{x}$ = 61.60 SD = 10.73 N = 10

As with raw scores, mental ages provide little information for between-group comparisons where age is not a constant factor. It is somewhat more helpful to collapse the scores above (Table 2-5) and represent them graphically as is done in Figure 2-2. Since mental age scores reflect national norms in themselves, no comparisons will be made with the normative sample by the very nature of the index.

TABLE 2-5

MENTAL AGE (IN MONTHS) MEANS, STANDARD DEVIATIONS, AND SAMPLE SIZES BY TREATMENT GROUPS

	Package	TV-HV	TV-only	Control
$\bar{x}$	57.97	55.94	50.36	52.69
SD	14.55	15.26	15.94	12.92
N	40	37	44	45

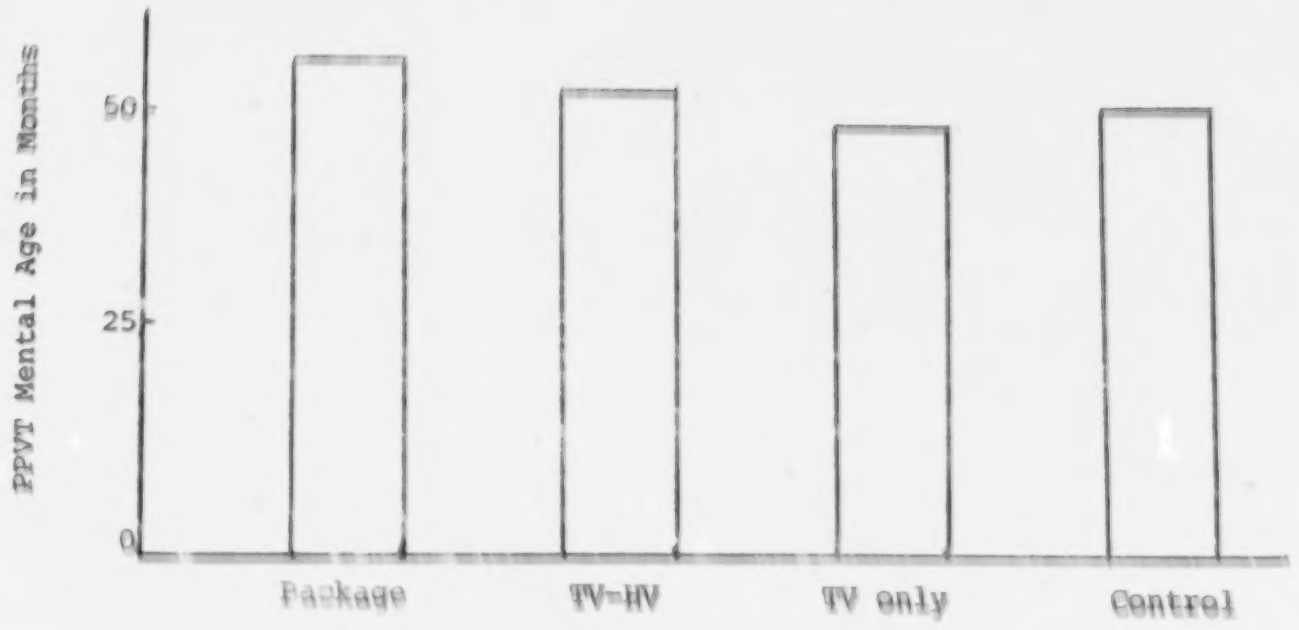


FIGURE 2-2

MEAN MENTAL AGE (IN MONTHS) FOR FOUR TREATMENT GROUPS

Although mental age scores are not derived by a precise formula, but rather are based on sample mean ages for a particular raw score, the analysis of variance summary below reflects the same treatment effect which was evident in the raw score analysis.

Also the main effect of age which was apparent throughout the entire test battery is also present in the mental age scores (Table 2-6), as was expected.

TABLE 2-6

ANALYSIS OF VARIANCE SUMMARY TABLE
FOR MENTAL AGE SCORES

Source	M^2	df	Mean Square	F	P
I(trt)	.033	3	393.3632554	2.02	
J(sex)	.002	1	79.90533782	0.41	
K(age)	.120	1	4331.799552	22.24	P .0005
IJ-INT	.000	3	4.846862524	0.02	
IK-INT	.004	3	45.09588620	0.23	
JK-INT	.001	1	37.70496729	0.19	
IJKINT	.028	3	333.8771177	1.71	
Error		150	194.7623583		

In the case of the mental age scores, as was true for group raw score means, the paraprofessional seemed to make a contribution to the level of vocabulary of the children whom she visited. Again, the TV only group produced the lowest score of the four samples, followed by the comparison group.

PPVT IQ

The IQ's which are derived from the PPVT are not numerical quotients, but are deviation figures based on the normative sample. For this reason, they may reflect the trends revealed by the raw scores, but the inexactness of the transformations may obscure some of the more subtle differences between groups. However, as contrasted with mental age, deviation IQ scores have the advantage of being independent of the child's age and provide a readily understood comparison with the individual's peer group.

Mean IQ scores for each age-by-sex subgroup within the four treatments are reported in Table 2-7.

TABLE 2-7

PPVT IQ SCORE MEANS, STANDARD DEVIATIONS,
AND NUMBER OF SUBJECTS BY AGE, SEX, AND TREATMENTS

Age	Sex	Package	TV-HV	TV only	Control
3	Male	$\bar{x}$ = 97.63 SD = 9.40 N = 8	$\bar{x}$ = 93.44 SD = 20.97 N = 9	$\bar{x}$ = 94.85 SD = 20.17 N = 13	$\bar{x}$ = 96.69 SD = 15.03 N = 13
	Female	$\bar{x}$ = 101.63 SD = 14.25 N = 8	$\bar{x}$ = 101.30 SD = 14.79 N = 10	$\bar{x}$ = 83.70 SD = 19.86 N = 10	$\bar{x}$ = 87.46 SD = 15.89 N = 13
4	Male	$\bar{x}$ = 99.38 SD = 12.20 N = 13	$\bar{x}$ = 102.50 SD = 8.11 N = 8	$\bar{x}$ = 88.63 SD = 23.08 N = 8	$\bar{x}$ = 92.33 SD = 15.78 N = 9
	Female	$\bar{x}$ = 94.81 SD = 17.33 N = 11	$\bar{x}$ = 95.30 SD = 16.67 N = 10	$\bar{x}$ = 88.38 SD = 32.73 N = 13	$\bar{x}$ = 93.90 SD = 16.24 N = 10

It is interesting to note that no consistent pattern of superiority for one sex is evident throughout the treatment groups or age subsets. Traditionally, girls are presumed to show increasing superiority in verbal development until adolescence. The Hooper & Marshall Pilot Study[3] also failed to show this superiority.

Combining these scores produces the results depicted graphically in Figure 2-3 and also in Table 2-8. Since these IQ scores imply a mean for each age (that of 100) no representation of the normative group is presented.

TABLE 2-8

IQ SCORE MEANS, STANDARD DEVIATIONS,
AND SAMPLE SIZES FOR FOUR TREATMENTS

	Package	TV-HV	TV-only	Control
$\bar{x}$	98.23	98.08	90.29	92.53
SD	13.64	15.79	20.29	15.58
N	40	37	44	45

[3]Frank H. Hooper and William H. Marshall, Final Report: The Initial Phase of a Preschool Curriculum Development Project, West Virginia University, Morgantown, W. Va. August, 1968.

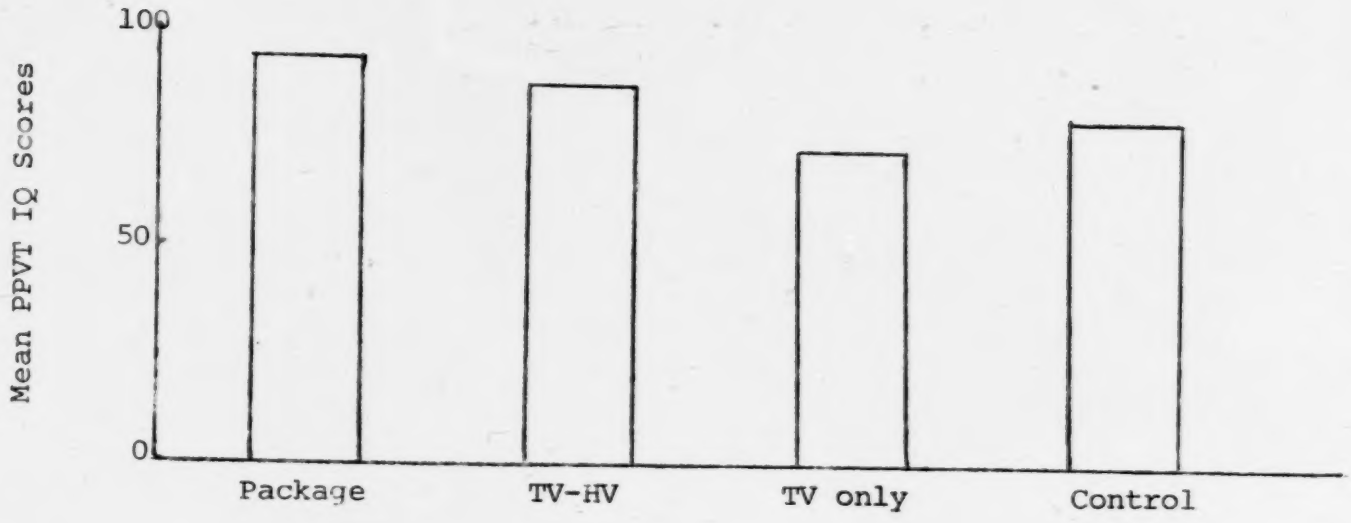


FIGURE 2-3

IQ SCORE MEANS FOR FOUR TREATMENT GROUPS

It is hypothesized that any difference which occurs between the control group and the treatment groups, and which favors the control group is caused by non-treatment factors. Since this is the case, it is likely that the TV only group has a slightly lower overall socio-economic status which is reflected in lower verbal ability. It is of interest that the control group which tended to score slightly above the TV only group, still produced means IQ's below the children in the "package" and TV-HV groups.

The analysis of variance summary table shown below in Table 2-9 further clarifies the results of the Peabody.

TABLE 2-9

SUMMARY OF ANALYSIS OF VARIANCE FOR PPVT IQ SCORES

Source	M^2	df	Mean Square	F	P
I(trt)	.047	3	722.1582158	2.56	
J(sex)	.003	1	150.3676385	0.53	
K(age)	.000	1	2.535774050	0.01	
IJ-INT	.003	3	51.71337161	0.18	
IK-INT	.002	3	33.59052361	0.12	
JK-INT	.000	1	1.489654994	0.00	
IJKINT	.033	3	514.7359580	1.82	
Error		150	282.2209675		

As can be seen from the above, no main or interaction effects are present at a statistically significant level. However, the fact that the Peabody IQ follows the same overall trend as the majority of the other subtests in this battery indicates that it is reflecting a similar distribution of ability and environmental effects.

SUMMARY AND CONCLUSIONS

Because of the highly specific nature of the test items on the Peabody, it is not likely that it reflects general program effects as well as the more broadly based instrument in a test battery.

In conclusion, two groups of children tested for the AEL Early Childhood Education Program (Package and TV-HV) scored near the national mean (50th percentile) in IQ and two groups (TV only and comparison group) scored near the 40th percentile when compared to the national sample. The lack of overall deficit indicates that many of the children of ages three and four in the AEL region have an adequate vocabulary level. Looking at raw score analysis, the results suggest the probability of a treatment effect in the verbal area which is reflected by the PPVT and which favors the Package and TV-HV groups.